



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

School of Computer Science and Statistics

Analysis of Media Sentiment Surrounding Brexit and Its Impact on the Irish Economy

Rory Goodman

14316993

April 12, 2019

A Dissertation submitted in partial fulfilment
of the requirements for the degree of
Masters in Computer Science

Declaration

I hereby declare that this project is entirely my own work and that it has not been submitted as an exercise for a degree at this or any other university.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at <http://www.tcd.ie/calendar>.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at <http://tcd-ie.libguides.com/plagiarism/ready-steady-write>.

Signed: _____

Date: _____

Abstract

Sentiment analysis is a type of natural language processing which infers emotion from text. Sentiment has been shown in previous studies to have an impact on stock markets.(1)(2) The United Kingdom's exit of the European Union is expected to have a significant impact on the Irish Economy(3) and therefore its stock market due to it being one of Ireland's major trading partners(4). What can sentiment analysis tell us about how the Irish Media feels about Brexit? How is that sentiment from the media affecting the Irish stock exchange?

A corpus of 18,000 articles from May 2018 to February 2019 was created using the LexisNexis media corpus. The corpus was created by taking any article with the term "brexit" in it. A domain specific Brexit dictionary was then created to use in addition to a general sentiment dictionary from Rocksteady. Negative sentiment was focused on in this study as it has been shown in previous studies(1) that pessimism in media is the main driver of sentiment's effect on markets. Sentiment was extracted from the corpus of about 18,000 articles in Irish publications and analyzed for negative sentiment using the Rocksteady Sentiment Analysis software. Different outputs from Rocksteady were then aggregated and analyzed using Pearson's correlation coefficient which p-values were generated from. Results from these analyses suggested general biases of negativity towards the UK politicians taking part in the negotiation compared to their Irish and European counterparts. It also showed significant differences in sentiment surrounding Brexit from different papers when compared. The Belfast Telegraph and Irish Independent were found to be significantly more pessimistic about Brexit compared to the Irish Times.

The sentiment data was then aggregated with pricing data from the ISEQ stock index, which is the national index of Ireland and is a good indicator of the overall economy. The data was then imported into GRET, a financial analysis program for regression analysis on time series. Regression analysis is a type of statistical modelling that analyzed the relationships between different variables in a time series. The analysis involved creating regression models of sentiment with market volume, volatility and returns. However, no statistically significant relationship was found between sentiment and any of these market indicators which was unexpected given the background research conducted.

Acknowledgements

I would like to thank my supervisor Professor Ahmad for all of his help with this study. I would also like to thank my parents for putting up with me these past couple of months and for all of their support throughout my time in Trinity.

Contents

1	Introduction	1
2	Background	3
2.1	Introduction	3
2.2	Asset Price Dynamics, Volatility and Prediction	3
2.3	Kumar, A., Lee, C. (2006)	3
2.4	Tetlock(2007)	4
2.5	Davies(2018)	5
2.6	Ahmad, Daly, Liston (2011)	5
2.7	How Sentimental are we about Brexit?	6
3	Methodology	8
3.1	Introduction	8
3.1.1	Goals	8
3.2	Tools	8
3.2.1	LexisNexis	10
3.2.2	Yahoo Finance	10
3.2.3	Rocksteady	10
3.2.4	Excel	11
3.2.5	GRETl	12
3.3	Implementation	13
3.3.1	Data Gathering	13
3.3.2	Sentiment Analysis	14
3.3.3	Financial Analysis	15
4	Results	18
4.1	Introduction	18
4.2	Sentiment Analysis	18
4.3	Financial Analysis	19
4.3.1	Returns Volume	20
4.3.2	Volatility	21

5 Conclusion	23
5.1 Interpretation of Results	23
5.2 Future Work	24
A1 Appendix	27
A1.1 Full GRETLE Outputs	27
A1.1.1 Daily Returns Negativity VAR	27
A1.1.2 Volume Negativity VAR	29
A1.1.3 Volatility Negativity VAR	30
A1.2 Brexit Dictionary	31

List of Figures

3.1	Layout of the data pipeline	9
3.2	LexisNexis Corpus	10
3.3	Rocksteady Affect Analysis Engine	11
3.4	Excel Spreadsheet	12
3.5	GRETLM	13
4.1	Daily Returns Negativity Scatter Plot	20
4.2	Volume Negativity Scatter Plot	21
4.3	Volatility Scatter Plot with Simple Least Squares Line of Best Fit	22

List of Tables

3.1	Table showing the top 10 sources by volume	14
4.1	Table showing comparisons of correlations of Negativity and Brexit Article Volume from different dictionary subcategories	19
4.2	Table showing comparisons of top 3 sources Rs	19

Nomenclature

EU	European Union
FTSE	Financial Times Stock Exchange
GRETLM	GNU Regression, Econometrics and Time-Series Library
GVC	Global Value Chain
ISEQ	Irish Stock Exchange Quotient
N	Sample Size
R	Pearson's Correlation Coefficient
TF-IDF	Term Frequency Inverse Document Frequency
UK	United Kingdom
VAR	Vector Autoregression

1 Introduction

This study aims to investigate the influence of media sentiment relating to Britain's decision to exit the European Union and the process stemming from it on the Irish economy. It was done using sentiment analysis software developed during a previous study at Trinity College.(5)

The United Kingdom is Ireland's 2nd largest importer(4) of goods and currently, before the UK's exit of the EU, both countries enjoy tariff free trade with each other due to their membership in the European Single Market. The absence of this tariff free trade in the event of a "no-deal" Brexit would likely have a large impact on the Irish Economy. The impact of a potential hard border between the Republic of Ireland and Northern Ireland in a "no deal" scenario could also negatively affect the Irish economy. This study aims to investigate media sentiment surrounding the Brexit process in Irish Publications and its relationship with the Irish economy.

For the purpose of this study we will be using daily prices from the ISEQ20 index as an indication of the reaction of the Irish economy. A stock index or stock market index is a measurement of a section of the stock market. It is computed from the prices of selected stocks, typically a weighted average. The ISEQ is an index which measures the top 20 stocks by trading volume on the Irish stock market. A 'national' index, such as the ISEQ, represents the performance of the stock market of a nation and gives a suggestion of investor sentiment on the state of its economy.

Sentiment analysis is a computational process which is used to measure positive, negative and neutral language in any type of source material. Also known as opinion mining, it can be used to measure people's attitudes or opinions on, for example, Twitter. In doing so, it can give us an early indication of broader public opinion. Political sentiment analysis can be used to determine how attitudes might change over a period of time. For example, Bloomberg carried a banner headline saying "Trump's Approval Rating Is Rising, and That's Bad for Stocks". The reporter showed that one of the economic barometers of US and possibly world economy, such as S and P 500, went down every time the US president's approval rating went up. It appears that people's sentiment has some impact on the movement of the US stock markets. ((6)).

This study aims to investigate whether a similar relationship exists between political sentiment and the Irish economy, particularly the recent sentiment surrounding the exiting of the United Kingdom from the European Union and the dynamics of the process stemming from it. The analysis will investigate effects on returns, volume and volatility of the ISEQ index.

This analysis continues from the analysis of sentiment surrounding Brexit and its impact on financial markets done by a team in the School of Computer Science and Statistics in Trinity College. The research took a large sample of articles from Irish publications including newspapers, magazines and web blog entries and observed a significant correlation between media sentiment surrounding Brexit and volatility in Irish markets. That study served as a guideline for many of the decisions made during this study.

The study used a number of tools in order to gather, aggregate and interpret both sentiment data and daily pricing data. These will be detailed in the methodology section.

2 Background

2.1 Introduction

This section details some of the background research that was conducted for this study. These are the sources which I found to be the most useful and significant in framing the context in which this study was conducted.

2.2 Asset Price Dynamics, Volatility and Prediction

Due to my lack of knowledge in both sentiment analysis and financial markets, there was a large amount of background research required in order to complete this project. I began by reading some relevant chapters of Stephen Taylor's *Asset Price Dynamics, Volatility and Prediction* in order to get to grips with some of the basics of daily prices and time series. The main takeaway from my reading of this book is that for analysing the effect of sentiment we should look at factors in terms of returns rather than prices due to the high correlation between daily prices. Other important factors which traders often consider are the volume of trading on a stock as well as the volatility of a stock. The volume of a stock indicates the number of trades on a stock over a given period. The volatility of a stock is the range of price change a security experiences over a given period of time.⁽⁷⁾

2.3 Kumar, A., Lee, C. (2006)

I then began looking into previous papers published which investigated the effects of sentiment on markets. One of the earliest studies into the impact of sentiment on markets was Kumar, A., Lee, C. (2006) "Retail investor sentiment and return co-movements" which was published in the *Journal of Finance*. This study aimed to challenge "traditional finance theory" which "posits that the current price of a stock closely reflects the present value of its future cash flows.". It instead aimed to establish the theory that the dynamic interplay

between noise traders, investors who make decisions regarding buy and sell trades without the support of professional advice or advanced fundamental analysis, and rational arbitrageurs establishes prices.

The analysis focused around individual investor sentiment as its noise trader model. The study showed that retail trades, also known as trades made by individual traders are heavily correlated. The trades were gathered from a popular retail trading house. They then used these trades as a guide for the retail sentiment surrounding a stock. The research they carried out on the markets found that in stocks with smaller market caps and lower trading volume were most affected by retail sentiment. They concluded that their results "support a role for investor sentiment in the study of financial markets". (2)

For the purposes of this study this establishes a dynamic between investor sentiment and markets to be studied.

2.4 Tetlock(2007)

Another relevant paper studied during my research was Tetlock(2007) "Giving Content to Investor Sentiment: The Role of Media in the Stock Market" also published in the Journal of Finance. It was the first major study to investigate the affect media sentiment can have on markets. The study set out to "quantitatively measure the interactions between the media and the stock market using daily content from a popular Wall Street Journal column."

The column, "Abreast of the market" gave a review of the market the previous day, whether it went "up, down or sideways" as measured by indexes such as the Dow Jones Industrial Average. The paper created a pessimism measure from the contents of the article and then used vector autoregression(VAR) to investigate the inter temporal relationship between the content of the Wall Street Journal and the stock market.

The results of the study showed that high degrees of media pessimism caused major downward pressure on markets which was then followed by a "return to fundamentals". However, the impact of high values of pessimism was shown to take significantly longer to return to the norm when it affected smaller stocks which is consistent with the work of Kumar. It also found weak correlation between pessimism and an increase in market volatility. It also found that the relationship between different measures of pessimism was the same i.e the reaction of the market to the prevalence of both Negative and Weak words. The study also posited that a sentiment based trading strategy with zero trading fees would be substantially profitable.(1)

2.5 Davies(2018)

Another study which investigated the influence on the market from external sources was "Davies(2018) The Heterogeneous Impact of Brexit: Early Indications from the FTSE" published in the European Economic Review. The study aimed to investigate the effect of the initial Brexit referendum result on June 23rd 2016 on the national index of the UK, the FTSE. The study took an approach of "examining how news of Brexit affected expectations as embodied in stock returns using a two-stage estimation process". The first stage of this was determining the initial impact of the referendum outcome on a firm's index via observing "the average deviation from the actual return relative to its predicted return based on the performance of the overall market". The second stage of the analysis involved investigating how a firm's performance following the referendum depended on the firm's global value chain structure. A global value chain in this context refers to the location of the firm's affiliates.

The results from the analysis showed firstly that the reaction was consistent with investors initially reacting to how the referendum could potentially impact on a firms' global value chains. Firms with their GVC's located heavily in Europe than ones with more globally or nationally oriented ones. The fall of the sterling following the referendum was expected to encourage exports and hamper imports. The study found that this impact was "contingent on the importance of a firm's imported intermediates.". The other key result was that the impact on the markets after the referendum was "swift and long lasting" for firms with affected GVCs. The initial losses of the firms affected persisted even as the market persisted from the initial shock. There was also little reaction from these firms to subsequent events surrounding Brexit such as the triggering of article 50. This suggests confidence from investors as to what Brexit means for these firms. The study established Brexit having a major, dynamic affect on markets which this study looks to investigate and expand upon.(8)

2.6 Ahmad, Daly, Liston (2011)

The study "What is new? News media, General Elections, Sentiment, and named entities" by Ahmad, Daly and Liston published by Trinity College from 2011 investigated the effect of sentiment in Irish publications surrounding the 2011 Irish general election. This study posited that the reuse of particular "affect" words for emphasis in publications surrounding the election could have an influence on the "electability" of some politicians. It aimed to investigate Irish publications for any biases towards particular parties or politicians and if any such endorsement had any impact on the outcome of the election. (5)

The researchers used the LexisNexis database of publications to gather articles to analyze. They decided to use any article from an Irish publication with "the terms Ireland and Politics, or Ireland and Elections(s) within the headline or opening paragraph.". It then analyzed the articles using the Rocksteady affect engine which is a piece of software developed at Trinity College for natural language processing. The software uses general language dictionaries, such as "Stone's General Inquirer Dictionary" in combination with more specific domain dictionaries. In this study, a specific domain dictionary was created: "The Irish general election dictionary contained candidate names as terms with party affiliation, constituency, party role, qualifications and gender as categories. This resulted in a dictionary with 517 terms and 39 categories". This was then used to infer relevancy of a piece of text, in this case articles using a frequency based approach rather than a more specific algorithmic one.

The results of the study showed that there was less bias towards the party in power, Fianna Fáil in 2011, than there was in the previous election where they retained power. It found that there was a large focus on the economic policies of each of the parties which coincides with a review of the structure of the Irish government's debt by the International Monetary Fund just before the election. It found significant gender bias towards male candidates, significantly more than the actual gender ratio of the candidates. The study inferred that the additional coverage the opposition parties of Labour and Fine Gael received suggested that these two parties would seize power which is what resulted from the election. The study displayed the effective use of the Rocksteady affect engine to generate meaningful sentiment analysis results and establishes the effective use of the system to analyze Irish media sentiment.(5)

2.7 How Sentimental are we about Brexit?

This article posted in RTÉ's freelance academic section "Brainstorm" detailed a brief study conducted by researchers at Trinity College's school of computer science and statistics. This study aimed to investigate the impact of sentiment surrounding Brexit on both the Irish stock market via the ISEQ, the Irish national stock market index, as well as the price of government bonds which tends to be a good indication of confidence of a country's population in its government.

The study used the LexisNexis corpus to gather a corpus of articles from various sources including The Financial Times, the Belfast Telegraph and a number of Irish publications including the Irish Times. The study then used Rocksteady to analyze the sentiment in these articles similar to the previous study mentioned. The main result of this study was the conclusion that the negativity surrounding Brexit appeared to have a substantial increase on

the volatility of the ISEQ. The volatility of a market is calculated by averaging the standard deviation of the distance of the price from the average price over a period of time. The study also found a less significant relationship between the ISEQ share price and Brexit negativity.

The study set a good precedent and basis for this study to continue with articles from May 2018 on wards, which is where the gathering of data for the published article ended and the gathering for this study begins. This study used similar methods to extend the articles' analysis up to March 2019 where the data gathering ended.(6)

3 Methodology

3.1 Introduction

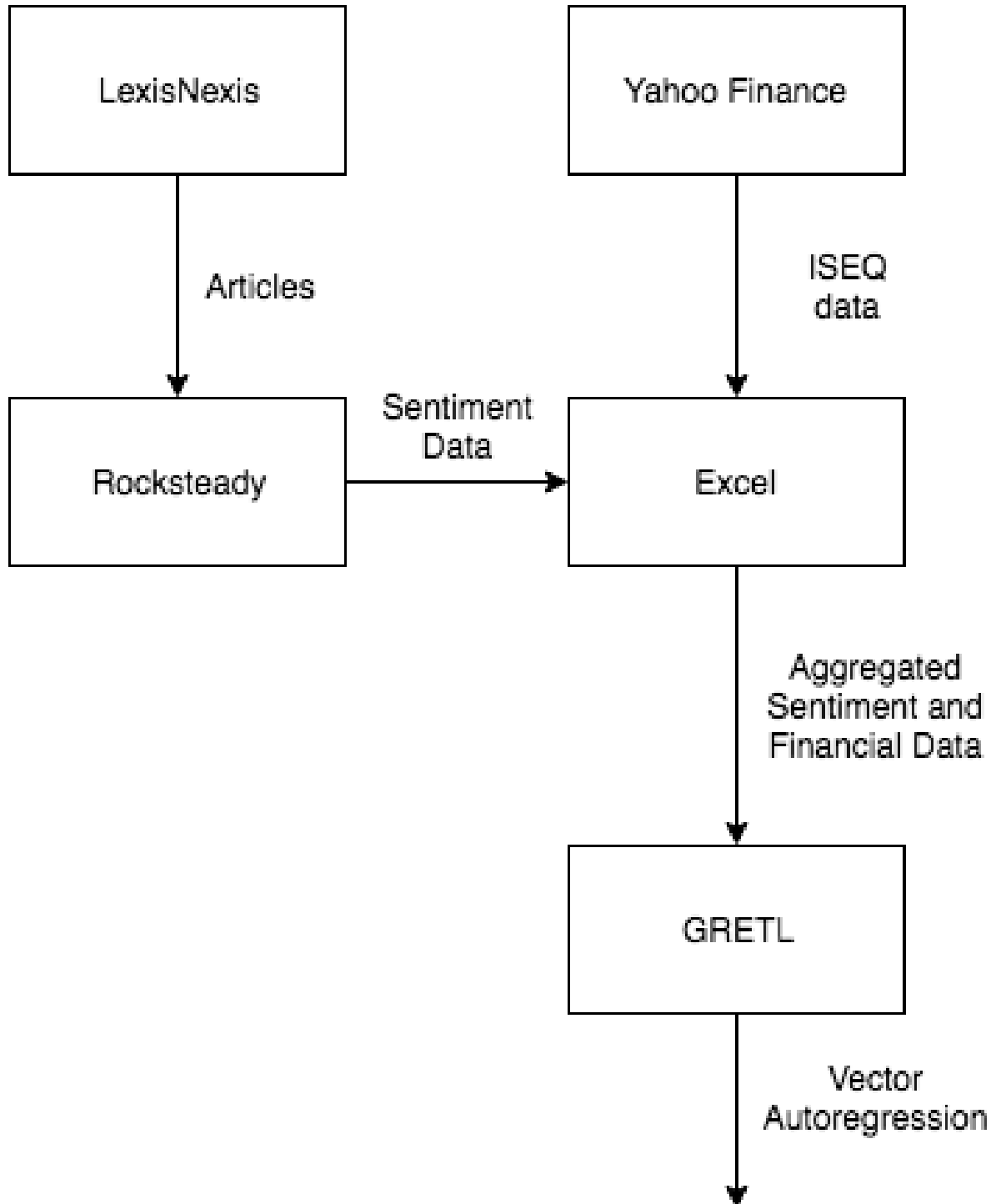
This section gives a detailed description of the research carried out. It details the goals of the research, the architecture and layouts of the tooling used for the research. It also will detail the reasoning and justification of the decisions taken in the layout of the system.

3.1.1 Goals

The goals of this study are to find what relationships exist, if any, between Ireland's national stock index the ISEQ and sentiment surrounding the exit of Ireland's largest trading partner Britain from the European Union in Irish Publications. This involves investigating whether changes in media sentiment are responsible for changes in returns, trading volume or volatility. It will also attempt to find any media biases surrounding Brexit from any publications in the articles gathered. These may include different levels of negativity between publications or varying sentiment in articles focused on one of the sides of the main negotiating parties in Brexit proceedings; the UK, the EU and Ireland.

3.2 Tools

Figure 3.1: Layout of the data pipeline



3.2.1 LexisNexis

To begin the project it was first necessary to gather a sufficient sample of relevant articles to analyze. For this I used the news and media corpus LexisNexis which allows for the gathering of articles from a huge number of publications. It allows for selection of results based on publications and a date range. It allows for different output formats of the articles. For this project, the articles were gathered in raw text files for compatibility with the other tooling used.

Figure 3.2: LexisNexis Corpus

The screenshot shows the LexisNexis search interface. At the top, there is a navigation bar with 'Nexis®', 'Search', and 'Documents'. On the right, it shows 'English (UK)' and the IP address '134.226.214.247'. Below the navigation bar, there is a search bar with 'News > Results: (brexit)' and an 'Edit' button. A yellow banner at the top of the results area states: 'Your search returned a large number of results. This list shows the first 1000 sorted by Relevance.' The main results area is titled 'Results (1000)' and includes a 'Select Language' dropdown and a 'Disclaimer' link. The results are sorted by 'Relevance'. The first five results are listed below:

Rank	Title	Source	Words	Hits	Date	Preview
1.	'What is' and 'when is' among most Googled questions on Brexit; The search engine results suggest voters are unsure about Britain's exit date.	Belfast Telegraph Online UK; Version:1,	232 words	22 hits	16 September 2018	Preview
2.	How political leaders reacted to the agreement on Brexit trade talks; The announcement came just before 7am.	Belfast Telegraph Online UK; Version:2,	610 words	20+ hits	08 December 2017	Preview
3.	Scrambled Brexit; LET TERS	Irish Daily Mail LETTERS; Pg. 39,	186 words	14 hits	21 March 2019	Preview
4.	The Ultimate Brexit Q and A; BRENDAN O'CONNOR	Sunday Independent NEWS; Pg. 1,	419 words	17 hits	24 March 2019	Preview
5.	Here's how Twitter reacted to a mammoth day in the world of Brexit; Resignations, a news conference and threats of a vote of no-confidence were all part of a busy day for the Prime Minister.					

3.2.2 Yahoo Finance

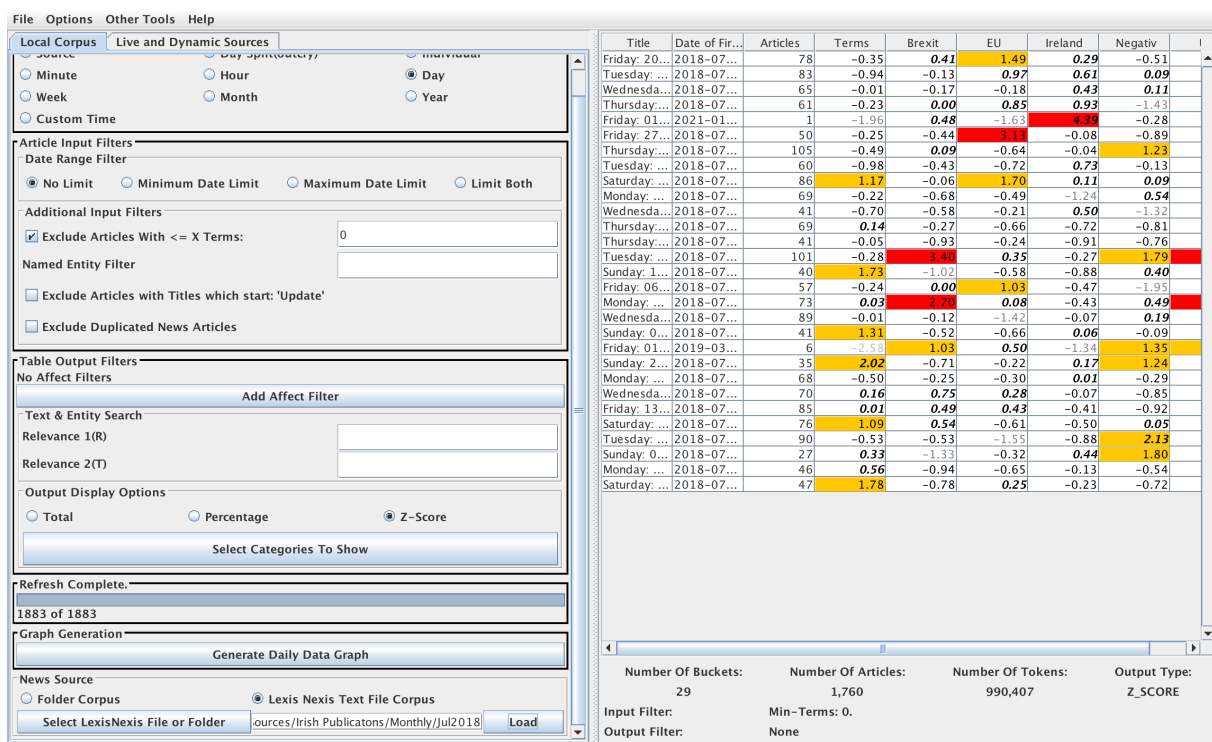
Yahoo! Finance is a media property that is part of Yahoo!'s network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. For the purposes of this project it was used to provide prices of the ISEQ for use in the regression analysis with the sentiment data from Rocksteady.(9)

3.2.3 Rocksteady

Rocksteady is an affect analysis system developed at Trinity College Dublin. According to Ahmad, Daly, & Liston, (2011) "The Rocksteady system uses a combination of general purpose affect dictionaries, like Stone's General Inquirer

Dictionary, and an optional domain specific dictionary” such as in the study previously mentioned where a dictionary of all of the candidates in the general election was analyzed with the general purpose dictionary. For this study a ‘Brexit’ domain specific dictionary was created and imported into Excel in order to add specificity to the results generated. This tool was chosen as it was shown to generate meaningful results in both the 2011 paper previously mentioned as well as the article from which this study aims to continue from. It was also chosen due to its compatibility with the LexisNexis corpus and for consistency across the study of 2018’s articles and the previously mentioned study of the 2017 sentiment impact.(6)

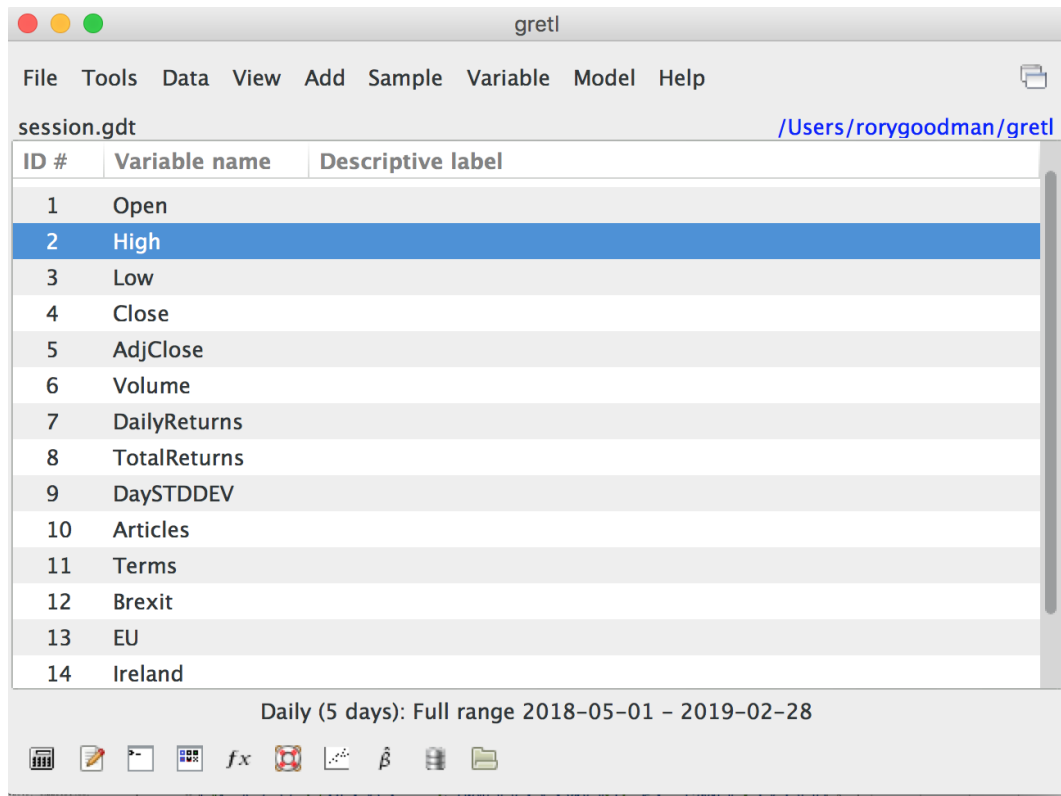
Figure 3.3: Rocksteady Affect Analysis Engine



3.2.4 Excel

Excel is a spreadsheet designed by Microsoft. According to Microsoft "Excel is an incredibly powerful tool for getting meaning out of vast amounts of data. But it also works really well for simple calculations and tracking almost any kind of information." It is a dynamic tool for the organization, graphing and interpretation of data and therefore is an ideal tool to use for the organization of the data output from Rocksteady as well as the financial data I used from Yahoo Finance.

Figure 3.5: GRET



3.3 Implementation

3.3.1 Data Gathering

The first part of the implementation of the project was to gather the articles needed to extract sentiment from. This was done by downloading all articles from Irish Publications which contained the word "Brexit" in them. This rule was chosen as it allowed for the gathering of relevant articles while still gathering a large enough sample size for the results extracted to be significant. The sources from which the articles were gathered were mostly from major newspapers such as the Belfast Telegraph, Irish Times and Irish Independent. A number of smaller publications such as tabloids, press releases and web publications were also included in the results. I chose to leave them in due to previous evidence of non traditional media sentiment influencing sentiment found by Bollen, Mao and Zeng(11).

Sources	No. of Articles
Belfast Telegraph Online	5491
The Irish Times	2793
Irish Independent	2469
Belfast Telegraph	1420
Irish News	1217
Irish Daily Mail	867
Sunday Independent	787
BreakingNews.ie	786
RTÉ News	757
Business World(Digest)	331

Table 3.1: Table showing the top 10 sources by volume

3.3.2 Sentiment Analysis

Before the articles were imported into Rocksteady for analysis, it was necessary to create a domain specific dictionary. Similar to the study on the Irish general election in 2011(5), where a dictionary of the candidates was created for analysis, a dictionary with relevant figures in Brexit negotiations was created. The selection of who to include was based heavily off of an article published on RTÉ's website in April 2018(12) just before the start date of the corpus used. Other names included were replacements of members in the dictionary who resigned during the period. One such example being the replacement of British Brexit Secretary David Davis by Dominic Raab in July 2018(13) and the subsequent replacement of Raab by Stephen Barclay(14). The dictionary was then broken down into people who were on the leave side of the referendum and people on the remain side of the referendum. The people were also divided into sub categories based on their side of the negotiations: Ireland, UK or the EU.

After this was imported into Rocksteady, I began to conduct analysis of the sentiment using the Negativity category from the General Language Dictionary in Rocksteady and the created Brexit dictionary. This study focused on negative sentiment due to Tetlock's finding of pessimism in media being the main driving factor of investor sentiment(1). First, I generated sentiment data to see if there was a correlation between the negativity and the number of articles posted on Brexit on a given day. The negativity figures were generated using a technique called Term Frequency - Inverse Document Frequency (TF-IDF). TF-IDF is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term has its respective TF and IDF score. The product of the TF and IDF scores of a term is called the TF-IDF weight of that term.(15). A z score is then calculated for each daily weight. A z-score is how many standard deviations a value is from the average of its sample. For the purposes of this study it will give a good

indication of the changes in sentiment throughout the ten month sample. Z-scores of TF-IDF were also used for the generation of relevance scores using the Brexit dictionary, these were used to determine the relevance to Brexit of articles as well as the relevance of the various parties in the Brexit dictionary for articles written on a given day.

The first stage of the analysis was to see if there were any trends or biases to be found in the sentiment data extracted from Rocksteady. I decided first to see if the number of Brexit articles posted in one day had an effect on the sentiment of the article. This was done by calculating the correlation coefficient of the daily series of articles and negativity z-scores which would give an indication whether the amount negativity of articles in the corpus was correlated with the volume of them. The correlation calculation was done with Excel which the daily time series of articles and negativity was exported to. The Excel correlation function was used for this which is based off of Pearson's correlation function "which measures the strength and direction of linear relationships between pairs of continuous variables."(16). The equation for its calculation is found below. I also investigated whether this value changed month to month over the sample to attempt to establish whether there were any significant variations in sentiment.

Pearson's Correlation Coefficient

$$r = \frac{\sum (x - \bar{X})(y - \bar{Y})}{\sqrt{[\sum (x - \bar{X})^2][\sum (y - \bar{Y})^2]}}$$

I then imported the sentiment data generated with the domain specific Brexit dictionary in an attempt to generate more meaningful results. I first calculated the correlation coefficient of the daily negativity and the overall Brexit relevance z-score. I also calculated the coefficient of negativity and relevance of the sub categories of the dictionary: Leave, Remain, UK, Ireland and EU to attempt to infer any leanings from the sample of articles. I also calculated these correlations on subsets of the samples restricted to a single publication. I chose to calculate these for the top three publications in the sample based on volume: The Belfast Telegraph, The Irish Times and the Irish Independent. From this it was possible to investigate whether any political leanings were prevalent in the sample.

3.3.3 Financial Analysis

For the financial portion of the analysis, first the daily ISEQ data was downloaded from Yahoo! Finance and imported into Excel. The Yahoo data provided 5 day daily time series for the ISEQ of the following data values: Open, High, Low, Close, Adjusted Close and

Volume. As this study aims to investigate sentiment's effect of Returns, Volume and Volatility, it was necessary to calculate the daily returns and volatility values from the other columns provided. Daily returns is calculated as the difference between the open and closing price of the index. Volatility is a rate at which the price of a security increases or decreases for a given set of returns. Volatility is measured by calculating the standard deviation of the returns over a given period of time. It shows the range to which the price of a security may increase or decrease. In this case the average volatility of the previous and next five trading days was averaged to give the 10 day volatility at that day. After these additional metrics were calculated for the financial data it was combined with the daily sentiment data so both time series could be imported into GRET. For GRET to import both the time series data and the index data it was necessary to reduce the sentiment data to a five day sample rather than a seven day. Therefore for the financial analysis, only articles published on trading days were considered. The data was then imported into GRET where scatter plots were created and Vector Autoregression was performed.

The main drawback of using correlation for these results is the lack of lagged variables. The use of VAR for the financial analysis solves this problem. Regression analysis is a statistical method for estimating the relationship among variables. Regression analyses how variables change over a sample and estimates the relationship by creating a line of best fit which indicates whether the relationship is positive or negative. R^2 is a calculation based on the mean distance of scatter plot points from the regression line. This value indicates how closely the regression line models the relationship. Specifically, for the purpose of this study Vector Autoregression (VAR) was used to examine the dataset as it allows for the model to take into account lagged variables of the value being analyzed. With values which are heavily influenced by their previous value particularly volume and volatility it is important that these are considered to improve the accuracy of the model. VAR is a model based on the plotting of endogenous and exogenous variables to a system. Endogenous variables are internal to a system. e.g Lagged values of returns. Exogenous variables are external to the system e.g sentiment. The VAR model has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting.(17). The basic principle when performing regressions is to create a scatter plot of two variables, in the case of this study returns and sentiment, volatility and sentiment and sentiment and trading volume. The next step is to attempt to fit a regression line to the data. The Adjusted Coefficient of Determination often referred to as the Adjusted R^2 describes the goodness of fit of the regression line to the data The full equation for calculating VAR used in this work is as follows, where v is the endogenous variable returns, volume or volatility, α and β represents the coefficients of 5 lags of return, s represents the coefficients of sentiment and ε the error term.

Vector Autoregression

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \beta_1 s_t + \beta_2 s_{t-1} + \varepsilon_t$$

4 Results

4.1 Introduction

This section details the outcomes of the study and attempts to infer meaning from them. Not all of the results were interesting or meaningful as is to be expected with this type of analysis. Therefore there will be significant focus on the more meaningful results in this section. Results' significance will be based on their p values. A p value of less than .05 indicates that a result is statistically significant and cannot be explained by random distribution. A p value of greater than .05 indicates that we accept the null hypothesis of no relationship between variables.

4.2 Sentiment Analysis

The values in this section are derived from Pearson's R correlation coefficient. The p values were calculated using Excel.

The analysis of the results output from Rocksteady in general were largely statistically significant. Many of the Pearson correlation coefficients generated using Excel on the data were below -0.1 and above 0.1 , which over a sample of 303 days is significant. It was found that the correlations of Brexit relevance and negative sentiment varied significantly month by month which would be expected given the large amount of new developments which occurred throughout the time span of the sample.(18)

The analysis of overall Brexit relevance and negative sentiment yielded an interesting result with a negative correlation between negative sentiment and Brexit relevance of -0.13 and a p for the sample of $.02$. This suggests that the articles with more relevance to Brexit had less negative terms in them. Another interesting result of this analysis was the variation in the correlation between negative sentiment between the sub categories of the domain specific Brexit dictionary. Over the sample an analysis was done of the different sub categories of Brexit negotiators in the Brexit dictionary: UK,EU and Irish politicians involved in the Brexit

deal negotiations. The correlation between the relevance of UK political figures in the Brexit dictionary were found to be significantly more correlated with negativity than those from the EU and Ireland.

Source	R	N	p val
UK	-.13.	303	.023626
EU	-.24	303	.000024
Ireland	-.22	303	.000113

Table 4.1: Table showing comparisons of correlations of Negativity and Brexit Article Volume from different dictionary subcategories

Perhaps the most interesting results from this portion of the analysis I believe were the variations on sentiment from the three major publications which were analysed. The study found a very significant variation in sentiment in Brexit articles between the Irish Times, Belfast Telegraph and Irish Independent. The correlation between the volume of Brexit articles and negative sentiment was statistically significantly positively correlated in the Belfast Telegraph. This is interesting considering the historically different political leanings would suggest that the Belfast Telegraph, a traditionally unionist paper(19), would display a more positive sentiment towards Brexit generally. It could be hypothesized that this may be indicative of their disappointment with the Brexit process throughout the sample, but more in depth analysis would be necessary to add merit to this hypothesis. The Irish Independent also showed statistically significant negativity in days where more Brexit articles were published. The Irish Times by comparison appeared to remain far more neutral in their reporting as their sample did not show any significant correlation between.

Source	R	N	p val
Belfast Telegraph	.28	303	<.0001
The Irish Times	-.05	303	.384207
Irish Independent	.18	303	< .00001

Table 4.2: Table showing comparisons of top 3 sources Rs

4.3 Financial Analysis

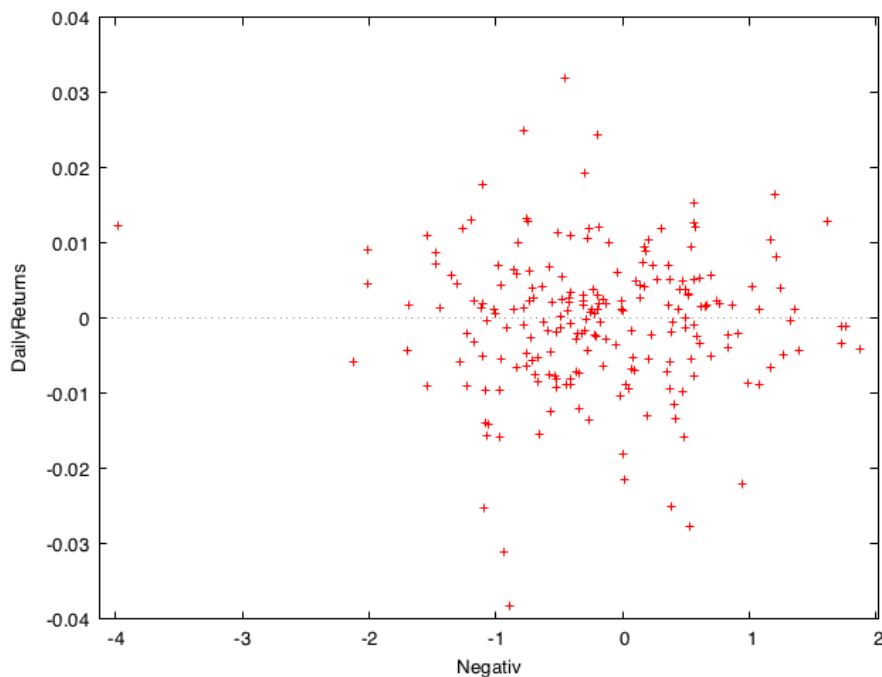
The regressions for returns, volume and volatility were set up with a lag of 5 for the ISEQ data. This meant that the 5 preceding days of the ISEQ value in each regression was considered with the actual daily value for each measure. It allows for the model to account

for the progression of the value in the proceeding days to increase the accuracy of the model. A scatter plot was also created for each relationship to improve readability of results. The VAR Analysis in GRETl returns a p-value of each variable in the VAR which indicates the statistical significance of the relationship between itself and the variables. When a p-value is less than .05 for a hypothesis we can reject the null hypothesis

4.3.1 Returns Volume

The first VAR performed with endogenous lagged variable Returns and exogenous variable returns over the period. This was done to attempt to find a statistically significant relationship between returns and sentiment, similar to Kumar's findings in 2006(2). The scatter plot of these two variables below.

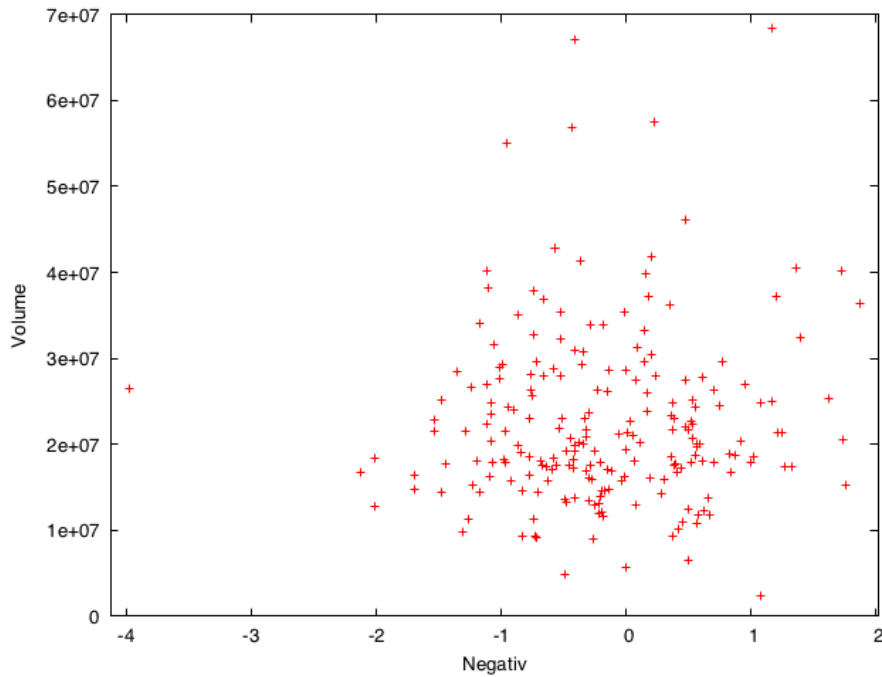
Figure 4.1: Daily Returns Negativity Scatter Plot



The scatter plot reflects the results shown in the VAR model. There was almost no suggestion that the returns were influenced by sentiment with a p value of .98.

A similar VAR was modelled with Volume values from the ISEQ. The scatter plot for these values is below.

Figure 4.2: Volume Negativity Scatter Plot

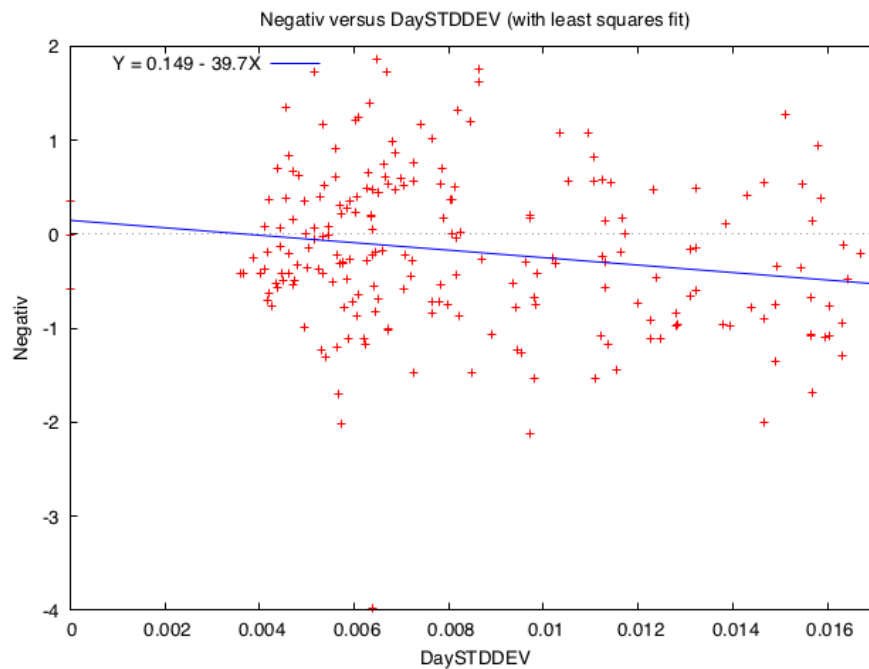


There was also no suggestion of significant relationship between Volume and negative media sentiment. This is consistent with the work of Tetlock(1).

4.3.2 Volatility

Volatility was measured as it was the measurement of the ISEQ which was found to be effected most by sentiment by both Tetlock and in the article the study builds upon.(6)(1) The final VAR was built with 10 day volatility as well as 5 day lagged values of them. Lagged values may have been unnecessary in this model as the volatility already takes into account surrounding days. It may have been possible to find similar results with a simple linear regression.

Figure 4.3: Volatility Scatter Plot with Simple Least Squares Line of Best Fit



The line of best fit was generally quite close to the data in the scatter graph. This is backed up by a high R squared value of .90 in the VAR results from GRET. However, the very high p-value returned suggests that the relationship between the two variables is not significant. This is inconsistent with the work of Tetlock which posited that Volatility was significantly influenced by media sentiment(1). It's possible that if more specific analyses were carried out that a relationship could have been found.

5 Conclusion

5.1 Interpretation of Results

The more interesting results from this study come from the analysis of the analysis of the sentiment and how the Irish media portrays Brexit in a broad scope. Days with more Brexit coverage had significantly less negative sentiment than the mean. Brexit articles with more mentions of the figures in the Brexit dictionary were less negative than those with less. This might be explained by articles detailing the ongoing negotiations being more matter of fact than other more opinion based articles. The analysis of the different subcategories is also interesting as it showed a substantial negative bias shown against the members of the UK politicians in the dictionary compared to the rest of the dictionary. The paper comparison also showed some significant biases between different publications of different publications. The Irish Independent, a conservative newspaper and The Belfast Telegraph, a Unionist newspaper showed statistically significant negative sentiment towards Brexit while the Irish Times a liberal newspaper showed no correlation between Brexit relevance and negative sentiment.

The regression analysis of the relationship between negative Brexit sentiment and the ISEQ showed that the negativity of Brexit articles had no significant relationship with returns or trading or volatility. This result is inconsistent with the work of Tetlock(1) and with that of Kumar(2). The result being inconsistent with Kumar was somewhat to be expected as it investigated based on the sentiment of retail investors rather than media sentiment. The inconsistency with Tetlock's work is more surprising as it did consider media sentiment but it is possible that the fact that its source of sentiment was from a popular column in a respected financial newspaper meant that it had more influence on investors. The most interesting takeaway from the financial results is that they are inconsistent with the ISEQ analysis of the year previous. The article did not specify a method of calculation of the results so it is possible that this could be explained by a use of different methods. It's possible that the potentially unnecessary use of VAR for analysis affected the result to the relationship being statistically insignificant. It is also possible that a miscalculation was made at some point in the pipeline leading to this unexpected result.

There are a number of limitations with this study, mainly the lack of event analysis during a time period where the dynamics of the ongoing Brexit negotiations changed drastically(18). Another shortcoming is the analysis of only the entire ISEQ index without analyzing individual companies' reactions to media sentiment. It would have been interesting to investigate whether the effects of companies with large volume of trading with the UK were affected differently from those who are mostly domestically based similar to the work of Davies who found there to be a discrepancy between companies with different GVCs.

5.2 Future Work

I believe that the findings in this study leave open opportunity for the continuation and expansion of this research in future projects. The simplest of which would be to extend this research past the cutoff date for the gathering of data. This would allow for the analysis of Brexit sentiment as the process continues past March 1st 2019. Even as of writing there have been several interesting developments in the Brexit negotiations process, such as the extension of the Brexit deadline from the initial date of the 29th of March 2019(20). It would also be nice to see a more in depth study of individual companies' prices in the ISEQ20. Perhaps it would be interesting to see if the work of Davies(8) holds true for the Irish stock market as well as the FTSE in that companies with a large part of their global value chain in Britain are significantly negatively affected by pessimism surrounding Brexit. Comparisons of companies such as PaddyPower Betfair who conduct a large part of their business in the UK and more domestically focused companies such as Bank of Ireland could certainly show some discrepancies in reaction to media Pessimism.

Another potential extension of this work would be to conduct an event analysis around key dates in the Brexit negotiations. The ISEQ hit its low point of the nine month sample in December of 2018(9) which was a month with key developments in the Brexit negotiations within the British parliament such as prime minister May's initial Brexit deal proposal being withdrawn due to strong backlash from the House of Commons.(21) It would be interesting to see if this coincidence would show any statistical significance in regression analysis. This concept could be expanded out to many other dates such as the extension of the Brexit date by the EU in March.(20)

Bibliography

- [1] Paul C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *SSRN Electronic Journal*, 2005. doi: 10.2139/ssrn.685145.
- [2] Alok Kumar and Charles M.c. Lee. Retail investor sentiment and return comovements. *SSRN Electronic Journal*, 2005. doi: 10.2139/ssrn.502843.
- [3] Noel Baker. New report reveals 'negative impacts of brexit. *Irish Examiner*. URL <https://www.irishexaminer.com/breakingnews/ireland/new-report-reveals-negative-impacts-of-brexit-on-irish-economy-in-all-scenarios-1.1111111.html>.
- [4] Top trading partners 2017 - cso - central statistics office, Dec 2018. URL <https://www.cso.ie/en/releasesandpublications/ep/p-ti/irelandstradinggoods2017/toptradingpartners2017/>.
- [5] Khurshid Ahmad, Nicholas Daly, and Vanessa Liston. What is new? news media, general elections, sentiment, and named entities. URL <https://aclweb.org/anthology/papers/W/W11/W11-3712/>.
- [6] Gormley-Heenan Kelly Finan Ahmad. How sentimental are we about brexit?, Apr 2018. URL <https://www.rte.ie/brainstorm/2018/0427/958625-how-sentimental-are-we-about-brexit/>.
- [7] Stephen Taylor. *Asset price dynamics, volatility, and prediction*. Princeton University Press, 2005.
- [8] Ronald B. Davies and Zuzanna Studnicka. The heterogeneous impact of brexit: Early indications from the ftse. *European Economic Review*, 110:1–17, 2018. doi: 10.1016/j.euroecorev.2018.08.003.
- [9] Yahoo finance - business finance, stock market, quotes, news. URL <https://finance.yahoo.com/>.

- [10] Giovanni Baiocchi and Walter Distaso. Gretl: Econometric software for the gnu generation. *Journal of Applied Econometrics*, 18(1):105–110, 2003. URL <https://EconPapers.repec.org/RePEc:jae:japmet:v:18:y:2003:i:1:p:105-110>.
- [11] Johan Bollen and Huina Mao. Twitter mood as a stock market predictor. *Computer*, 44(10):91–94, 2011. doi: 10.1109/mc.2011.323.
- [12] RTÉ News. Brexit: The key players, Mar 2018. URL <https://www.rte.ie/news/brexit/2018/0329/950824-brexit-five-key-players/>.
- [13] Jessica Elgot and Dan Sabbagh. Dominic raab named brexit secretary in cabinet reshuffle, Jul 2018. URL <https://www.theguardian.com/politics/2018/jul/09/dominic-raab-appointed-new-brexit-secretary-in-uk-cabinet-reshuffle>.
- [14] Stephen barclay named new brexit secretary, Nov 2018. URL <https://www.bbc.com/news/uk-46241693>.
- [15] Jurij Leskovec, Anand Rajaraman, and Jeffrey David. Ullman. *Mining of massive datasets*. Cambridge University Press, 2016.
- [16] Spss tutorials: Pearson correlation. URL <https://libguides.library.kent.edu/spss/pearsoncorr>.
- [17] Vector autoregressive models for multivariate time series. URL <https://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf>.
- [18] Ft collection: Brexit 2018, Jan 2019. URL <https://www.ft.com/content/72772246-fc99-11e8-ac00-57a2a826423e>.
- [19] Brendan Mcgrath. Independent buys 'belfast telegraph' for £300m, Mar 2000. URL <https://www.irishtimes.com/business/independent-buys-belfast-telegraph-for-300m-1.257174>.
- [20] Brexit: Departure date pushed back by at least two weeks, Mar 2019. URL <https://www.bbc.com/news/uk-politics-47663031>.
- [21] Dan Sabbagh and Jessica Elgot. Theresa may postpones brexit deal vote, Dec 2018. URL <https://www.theguardian.com/politics/2018/dec/10/theresa-may-postpones-brexit-deal-meaningful-vote-eu>.

A1 Appendix

A1.1 Full GRETLE Outputs

A1.1.1 Daily Returns Negativity VAR

VAR system, lag order 5

OLS estimates, observations 2018-05-09–2019-02-28 ($T = 208$)

Log-likelihood = 681.804

Determinant of covariance matrix = $8.32478e-05$

AIC = -6.4885

BIC = -6.3762

HQC = -6.4431

Portmanteau test: $LB(48) = 37.2027$, $df = 43$ [0.7200]

Equation 1: DailyReturns

	Coefficient	Std. Error	<i>t</i> -ratio	p-value
const	-0.000535832	0.000663014	-0.8082	0.4199
DailyReturns_1	-0.0685222	0.0703733	-0.9737	0.3314
DailyReturns_2	0.140711	0.0699177	2.013	0.0455
DailyReturns_3	-0.0185968	0.0705704	-0.2635	0.7924
DailyReturns_4	-0.0733213	0.0698640	-1.049	0.2952
DailyReturns_5	0.0532263	0.0698727	0.7618	0.4471
Negativ	0.000145996	0.000820873	0.1779	0.8590

Mean dependent var	-0.000571	S.D. dependent var	0.009303
Sum squared resid	0.017316	S.E. of regression	0.009282
R^2	0.033470	Adjusted R^2	0.004618
$F(6, 201)$	1.160077	P-value(F)	0.329220
$\hat{\rho}$	-0.006167	Durbin-Watson	2.007941

F-tests of zero restrictions

All lags of DailyReturns	$F(5, 201) = 1.36474$	[0.2391]
All vars, lag 5	$F(1, 201) = 0.580279$	[0.4471]

For the system as a whole —

Null hypothesis: the longest lag is 4

Alternative hypothesis: the longest lag is 5

Likelihood ratio test: $\chi_1^2 = 0.600$ [0.4387]

A1.1.2 Volume Negativity VAR

VAR system, lag order 5

OLS estimates, observations 2018-05-09–2019-02-28 ($T = 208$)

Log-likelihood = -3643.55

Determinant of covariance matrix = $9.60817e+13$

AIC = 35.1014

BIC = 35.2137

HQC = 35.1468

Portmanteau test: $LB(48) = 41.9461$, $df = 43$ [0.5169]

Equation 1: Volume

	Coefficient	Std. Error	t-ratio	p-value
const	1.86963e+07	3.18038e+06	5.879	0.0000
Volume _{t-1}	0.175912	0.0706779	2.489	0.0136
Volume _{t-2}	0.0777976	0.0717937	1.084	0.2798
Volume _{t-3}	-0.116222	0.0713240	-1.629	0.1048
Volume _{t-4}	-0.0170783	0.0716355	-0.2384	0.8118
Volume _{t-5}	0.0592874	0.0705042	0.8409	0.4014
Negativ	658604.	883152.	0.7457	0.4567

Mean dependent var	22643261	S.D. dependent var	10099572
Sum squared resid	2.00e+16	S.E. of regression	9971353
R^2	0.053484	Adjusted R^2	0.025230
$F(6, 201)$	1.892955	P-value(F)	0.083629
$\hat{\rho}$	-0.003318	Durbin-Watson	1.997079

F-tests of zero restrictions

All lags of Volume	$F(5, 201) = 2.1302$	[0.0633]
All vars, lag 5	$F(1, 201) = 0.707125$	[0.4014]

For the system as a whole —

Null hypothesis: the longest lag is 4

Alternative hypothesis: the longest lag is 5

Likelihood ratio test: $\chi_1^2 = 0.730$ [0.3927]

A1.1.3 Volatility Negativity VAR

VAR system, lag order 5

OLS estimates, observations 2018-05-09–2019-02-28 ($T = 208$)

Log-likelihood = 1116.54

Determinant of covariance matrix = 1.27343e-06

AIC = -10.6686

BIC = -10.5563

HQC = -10.6232

Portmanteau test: LB(48) = 92.3038, df = 43 [0.0000]

Equation 1: DaySTDDEV

	Coefficient	Std. Error	t-ratio	p-value
const	0.000496975	0.000212108	2.343	0.0201
DaySTDDEV _{t-1}	0.944972	0.0703334	13.44	0.0000
DaySTDDEV _{t-2}	0.118996	0.0940473	1.265	0.2072
DaySTDDEV _{t-3}	-0.0870021	0.0910960	-0.9551	0.3407
DaySTDDEV _{t-4}	0.0346354	0.0916547	0.3779	0.7059
DaySTDDEV _{t-5}	-0.0716831	0.0659577	-1.087	0.2784
Negativ	-2.09193e-06	0.000105955	-0.01974	0.9843
Mean dependent var	0.008511	S.D. dependent var	0.003671	
Sum squared resid	0.000265	S.E. of regression	0.001148	
R^2	0.905029	Adjusted R^2	0.902194	
$F(6, 201)$	319.2404	P-value(F)	7.6e-100	
$\hat{\rho}$	-0.006066	Durbin-Watson	2.008729	

F-tests of zero restrictions

All lags of DaySTDDEV	$F(5, 201) = 364.375$	[0.0000]
All vars, lag 5	$F(1, 201) = 1.18114$	[0.2784]

For the system as a whole —

Null hypothesis: the longest lag is 4

Alternative hypothesis: the longest lag is 5

Likelihood ratio test: $\chi_1^2 = 1.219$ [0.2696]

A1.2 Brexit Dictionary

Brexit	Leave	Remain	Ireland	EU	UK
May	Brexit		Remain		UK
Davis	Brexit	Leave			UK
Johnson	Brexit	Leave			UK
Rees-Mogg	Brexit	Leave			UK
Foster	Brexit	Leave		Ireland	UK
Barnier	Brexit		Remain		EU
Juncker	Brexit		Remain		EU
Verhofstadt	Brexit		Remain		EU
Tusk	Brexit		Remain		EU
Varadkar	Brexit		Remain	Ireland	
Coveney	Brexit		Remain	Ireland	
Raab	Brexit	Leave			UK
Corbyn	Brexit		Remain		UK
Cooper	Brexit		Remain		UK
Merkel	Brexit				EU